Metabarcoding and Metagenomics 5: 51–57
DOI 10.3897/mbmg.5.66503

Methods

Potential for cross-contamination of diatom DNA samples when using toothbrushes

Martyn G. Kelly[1,2], **Tim Jones**[3], **Kerry Walsh**[4]

1 *Bowburn Consultancy, 11 Monteigne Drive, Bowburn, Durham DH6 5QB, UK*
2 *School of Geography, University of Nottingham, Nottingham NG7 2RD, UK*
3 *Environment Agency, Sunrise Business Park, Higher Shaftesbury Road, Blandford Forum DT11 8ST, UK*
4 *Environment Agency, Horizon House, Deanery Road, Bristol BS1 5AH, UK*

Corresponding author: Martyn G. Kelly (mgkelly@bowburn-consultancy.co.uk)

Academic editor: Emre Keskin | **Received** 25 March 2021 | **Accepted** 6 May 2021 | **Published** 17 May 2021

Abstract

The use of toothbrushes and similar devices for sampling diatoms from hard surfaces is a well-established approach. Toothbrushes are routinely cleaned and reused when sampling for analysis by light microscopy. This paper looks at the scale of contamination encountered when this technique is used to sample diatoms for metabarcoding analyses, as well as at the scale of contamination to be expected if stream, rather than distilled water, is used to wash diatoms from stones. Although some contamination attributable to toothbrushes was detected, read numbers were low and had no effect on index calculation or ecological status estimates. However, if the primary focus of a study is to thoroughly document diversity in a sample, then even this small level of contamination may be unacceptable and more stringent measures may be required.

Key Words

diatoms, metabarcoding, ecological assessment, Water Framework Directive

Introduction

Although DNA-based technologies have potential to improve ecological assessment (Hänfling et al. 2016; Bista et al. 2017, 2018; Blackman 2017; Valentin et al. 2019; Kelly et al. 2020, Di Muri et al. 2020), regulators are still concerned about these technologies, particularly when they are replacing established methods. Challenges associated with metabarcoding methods include differences in the mode of quantification compared to traditional approaches (Kelly et al. 2020; Vasselon et al. 2018), in optimising bioinformatics pipelines (Baillet et al. 2020), in curating barcode libraries (Rimet et al. 2019, 2021) and in the lack of robust models to quantify uncertainties in DNA-based workflows. These technical challenges are being addressed by the research community, but some of the more subtle challenges and bottlenecks to implementing these technologies are less widely-recognised. These include logistical challenges associated with upscaling these new methods into large scale monitoring programmes and how they can be integrated into existing organisation infrastructures. Whilst there is a need to strive for the most robust, scientifically credible method, this must be balanced against a need from the user community for pragmatic, sustainable and cost-effective methods. If we fail to recognise this, it will impede uptake of methods by the end-user communities.

The standard method for sampling diatoms for ecological status and water quality assessment in Europe and beyond is to brush or scrape the upper surface of a hard substratum (Kelly et al. 1998, CEN 2014; Charles et al. 2020). Most workers use a toothbrush for this purpose, often reusing the same toothbrush at several sites and using stream water to rinse the biofilm off the stones and into containers. Kelly and Zgrundo (2013) showed that the scale of contamination associated with this approach

Copyright *Martyn G. Kelly et al.* This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

was low and was unlikely to have a significant effect on ecological status assessments when diatoms were analysed by light microscopy. By contrast, sampling for molecular ecology studies typically uses disposable, sterile equipment (see, for example, Bista et al. 2017). However, such approaches generate large quantities of non-biodegradable waste or require the transportation and use of environmentally unfriendly chemicals (e.g. bleach) for sterilisation in the field, as well as requiring samplers to carry pure water in the field. The issue of non-biodegradable plastic waste and disposal of used sterilising solutions becomes a more significant issue when sampling is scaled up for nationwide assessment programmes. Asking whether such strict attention to contamination is as relevant when sampling phytobenthos communities as when sampling water for eDNA, therefore, has a number of potential benefits, including lower cost and reduced waste.

This study investigates the scale of contamination introduced by the reuse of toothbrushes at different sites and on the effect of using stream water, rather than pure water, to rinse biofilms from surfaces. It has the same basic design as the study by Kelly and Zgrundo (2013) on contamination picked up during light microscopic studies, with two locations with very different ecological profiles being sampled with toothbrushes, some of which were previously unused and some of which had been used already at the other site. As the two sites have very few diatom species in common, this design should make it easy to pick out any contamination.

Materials and methods

Study design

Details of the sites are given in Table 1. Both are located in southern England: the River Nadder is a chalk stream in Wiltshire which is classified as having moderate ecological status, with macrophytes and phosphorus driving the classification (fish are at good status, invertebrates are high status and other chemical parameters are all high status). Ober Water, in contrast, is a stream in the New Forest (Hampshire) with softer (but still circumneutral) water and which is classified as being at good ecological status, with macrophytes and phytobenthos and all chemical parameters at high status. Further information on both sites can be found at: https://environment.data.gov.uk/catchment-planning/, https://environment.data.gov.uk/ecology-fish/ and https://environment.data.gov.uk/water-quality/view/landing.

Table 1. Background information on the sites used in the study. Values for chemical variables are medians (minimum – maximum) from the closest water chemistry sampling point for the 12 months before March 2017 (R. Nadder: 5 samples) and for 2015, the most recent data for Ober Water prior to sampling (6 samples).

	Nadder, Tisbury Station	Ober Water, upstream A35
Location:		
UK National Grid Reference	ST 94616 29152	SU 24964 03815
Latitude/Longitude	51°03'N, 002°04'W	50°50'N, 001°38'W
Altitude (m)	90	30
Water chemistry:		
Alkalinity (mgl^{-1} $CaCO_3$)	186 (82–205)	16 (10–20)
pH	8.0 (7.8–8.2)	8.1 (7.5–8.6)
Conductivity (μScm^{-1})	512 (246–527)	142 (120–147)
Ammonia-N (mgl^{-1})	0.042 (0.03–0.157)	< 0.03–0.063
Nitrate-N (mgl^{-1})	4.12 (1.98–4.45)	< 0.196–0.246
Reactive P (mgl^{-1})	0.169	0.0047 (0.0031–0.0068)5
Current ecological status:		
Overall	Moderate	Good
Macrophytes and phytobenthos	Moderate	High
Phosphorus	Moderate	High

Five samples were collected at each site for each of the following three treatments:

- brand new toothbrushes using distilled water to wash the biofilm into sample bottles;
- brand new toothbrushes using stream water from site to wash the biofilm into sample bottles; and
- toothbrushes previously used at the other site, along with stream water from the sampling site

In addition, three control samples (5 ml each) were collected:

- one using just distilled water; and,
- one each using river water from the two sites.

Sampling and analysis of benthic diatoms

Sampling involved brushing the upper surface of five cobble-sized stones and collecting the suspension in a tray. Using a new, but non-sterile Pasteur pipette, 5 ml of the suspension of biofilm and water was transferred to a sterile 15 ml centrifuge tube containing 5 ml nucleic acid preservative, consisting of 3.5 M ammonium sulphate, 17 mM sodium citrate and 13 mM ethylenediaminetetraacetic acid (EDTA). Samples were then transferred to the laboratory in a cool box and frozen at -30 °C prior to DNA extraction. The methods used for DNA extraction, amplification and analysis followed methods described in Kelly et al. (2020).

Data analysis

Non-metric multidimensional scaling (NMDS: McCune and Grace 2002) was used to investigate the structure of metabarcoding datasets using the vegan package in the R software package (R Core Team 2017; Oksanen et al. 2007) for multivariate analyses.

The Trophic Diatom Index (TDI5NGS) was calculated following Kelly et al. (2020) using the R package DARLEQ3 available at https://github.com/nsj3/darleq3. When evaluating the scale of variation in TDI5NGS, we used data from the Environment Agency (2018) which showed the average level of variation measured at a site on a single day. Kruskal-Wallis tests were implemented using base functions in R.

Results

The distilled water control sample contained just 23 reads compared with an average of 27,181 reads for all other samples. This control sample is not considered further.

The diatom assemblage from the River Nadder in samples collected with unused toothbrushes was dominated by *Navicula lanceolata* (average relative abundance: 41%) along with *Amphora pediculus* (8%), *Melosira varians* (8%), *Nitzschia recta* (7%) and *Navicula gregaria* (5%). In contrast, the diatom assemblage from Ober Water was dominated by *Achnanthidium minutissimum* (38%), along with *Platessa oblongella* (15%) and *Gomphonema truncatum* (11%).

Non-metric multidimensional scaling (NMDS) of the dataset yielded an ordination with very low stress (0.0684), with a clear separation between the two sites along axis 1 (Figure 1). However, some samples from

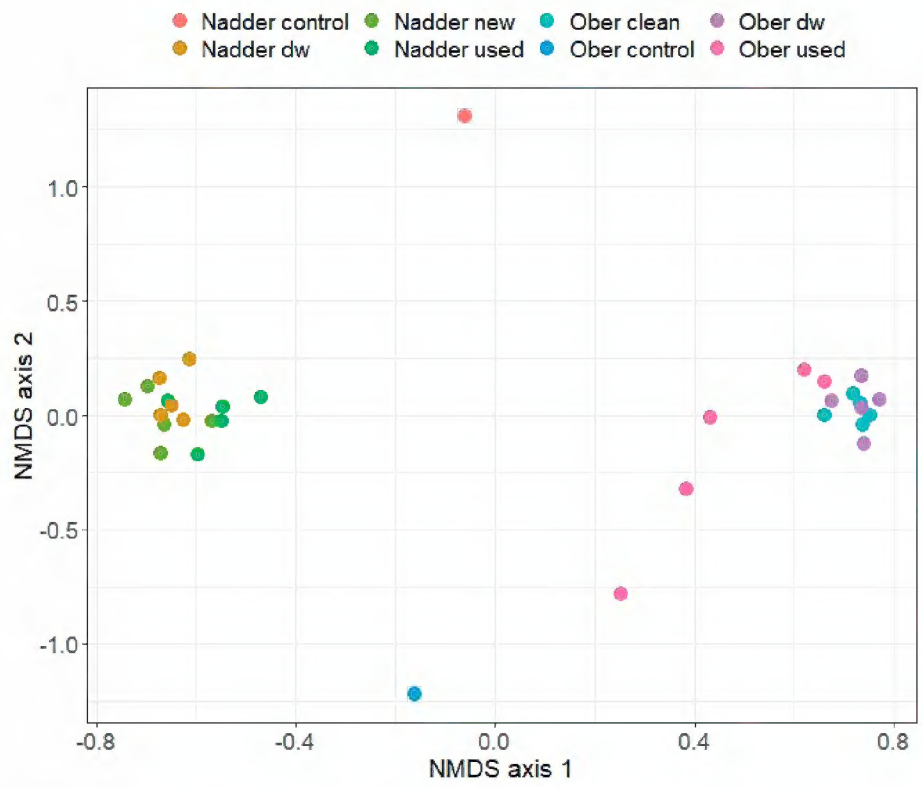

Figure 1. Plot showing position of samples from River Nadder and Ober Water relative to the first 2 axes of an NMDS ordination.

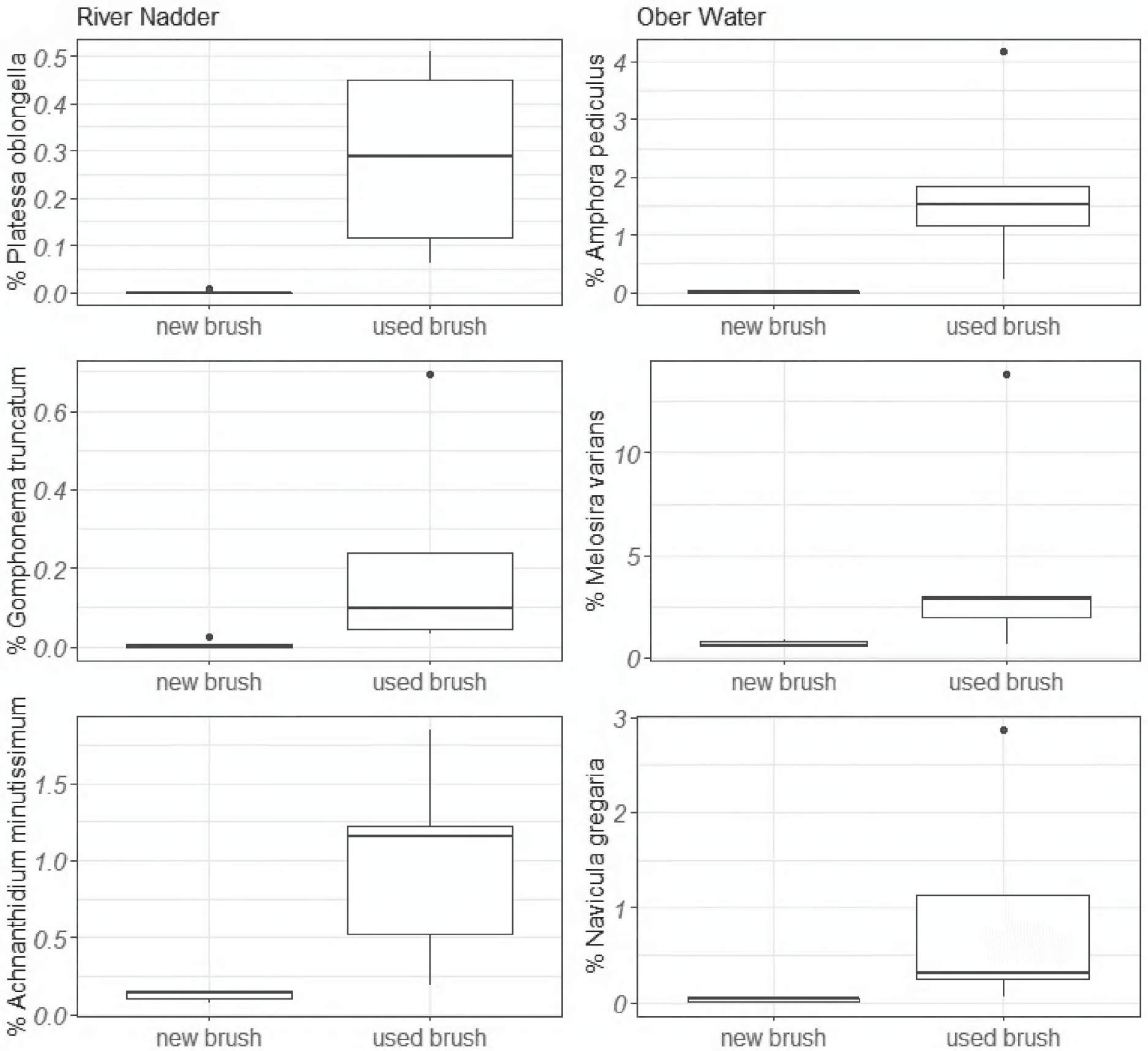

Figure 2. Variation in relative abundance of taxa carried over from one site to the other, using a clean toothbrush and one used previously at the other site.

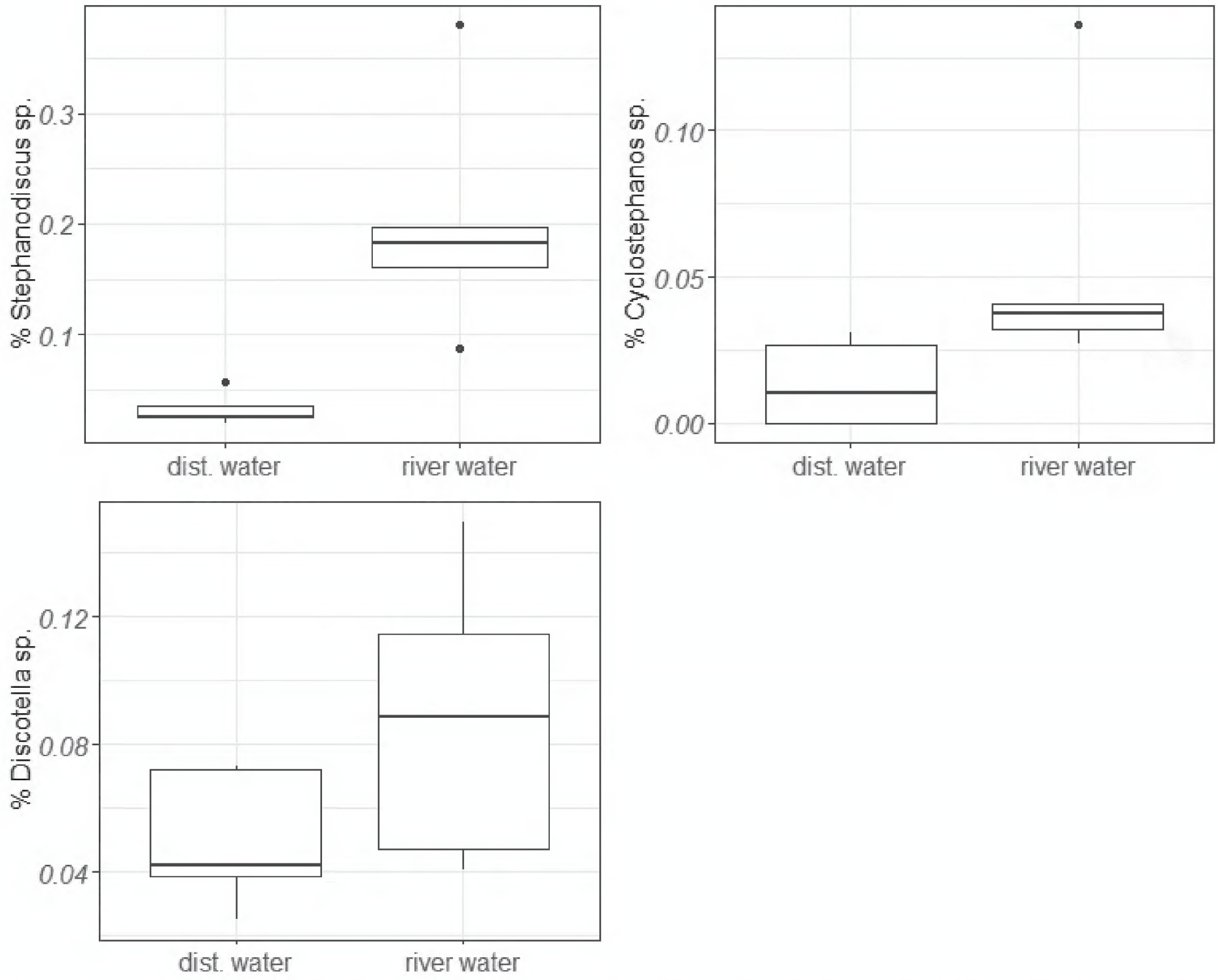

Figure 3. Variation in proportions of taxa that were abundant in stream water from the River Nadder at the time of sampling in biofilm samples collected with stream and distilled water, respectively.

Ober Water which were scrubbed using toothbrushes previously used at the Nadder site ("Ober used") had lower scores on the axis 1 than those scrubbed with clean toothbrushes ("Ober new"), suggesting some contamination from the River Nadder. By contrast, there was very little difference in the positions of samples collected using unused toothbrushes ("Nadder new") and toothbrushes previously used to sample the Ober ("Nadder used"). There was also very little difference in the position of samples rinsed with distilled water ("Nadder dw", "Ober dw") rather than river water ("Nadder clean", "Ober clean"). The river water control samples ("Nadder control" and "Ober control" in Figure 1) are distinct both from each other and from the biofilm samples along axis 2.

If there is a significant amount of contamination, then taxa that are abundant at one site should be present in raised numbers in samples collected using dirty equipment at the other site, but rare in the others. Although significant effects were observed for several taxa, the scale of the effect was generally small, particularly for samples from the River Nadder where the increased representation in samples collected with contaminated toothbrushes exceeded 1% only for *Achnanthidium minutissimum* (Figure 2). The scale of the increase was greater in Ober Water samples, with a median increase for *Melosira varians* of about 2%, but with one replicate having an increase >10% relative to the sample collected with clean equipment.

A similar approach was adopted to look at possible contamination from stream water. The relative abundance of the most abundant taxa in the stream water sample from each stream was compared with the samples washed with stream and distilled water from that location.

In the case of the River Nadder, the stream water was dominated by planktonic diatoms (65% of total reads). Three of these – *Stephanodiscus hantzschii*, *Cyclostephanos invisitatus* and *Discotella* sp. – were all elevated with respect to the distilled water sample (Figure 3), but only in relatively small numbers (that is, still < 1% in the worst case). Differences between treatments for *S. hantzschii* and *C. invisitatus* were both significant (Kruskal-Wallis tests: $p = 0.009$ and 0.016, respectively).

There were almost no planktonic diatoms in the Ober Water stream water; however, the composition of the sample was quite different to that of biofilm samples, with a greater proportion of nutrient-rich taxa. There was, despite this, no significant increase in proportions of these

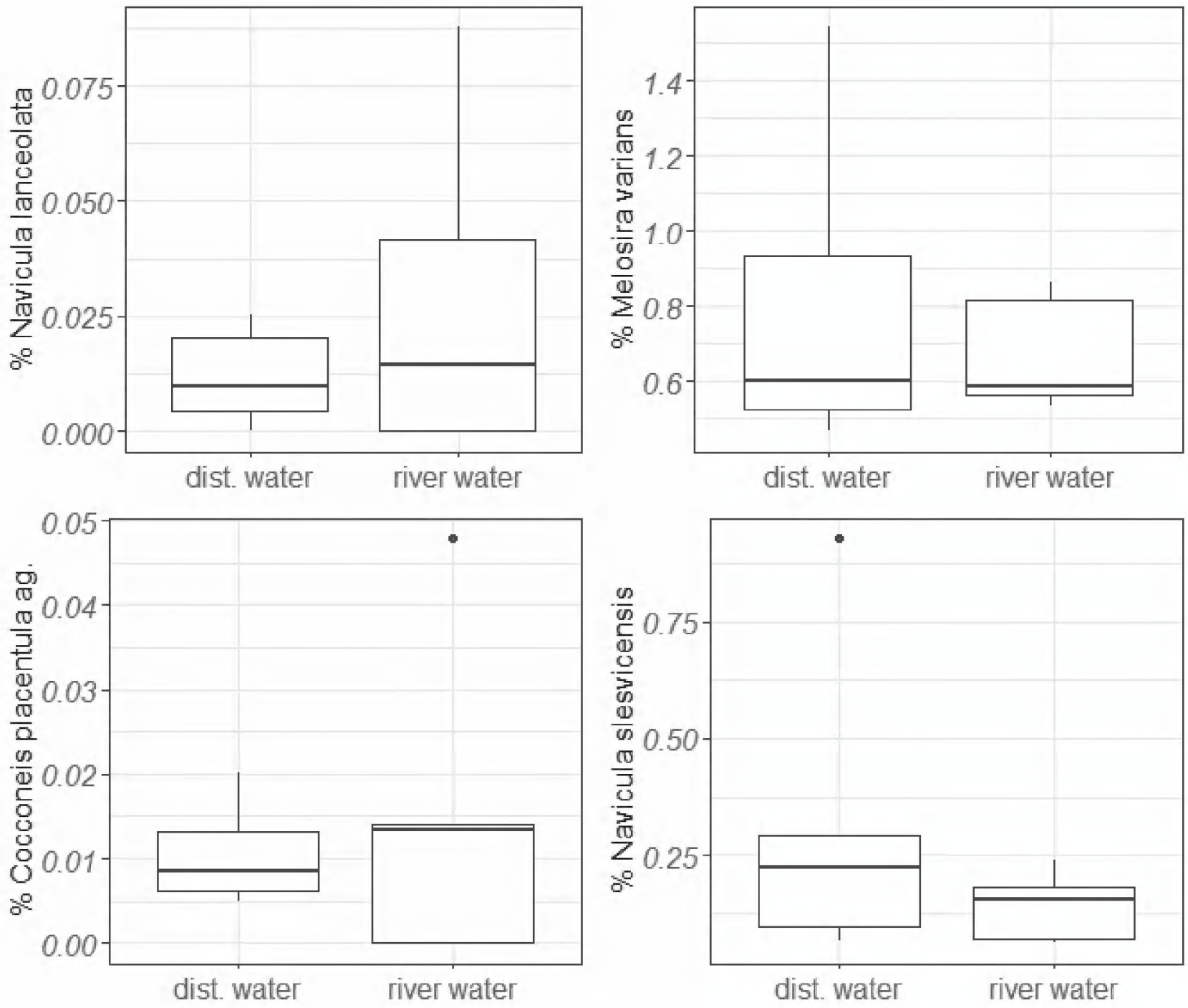

Figure 4. Variation in proportions of taxa that were abundant in stream water from Ober Water at the time of sampling in biofilm samples collected with distilled and river water, respectively.

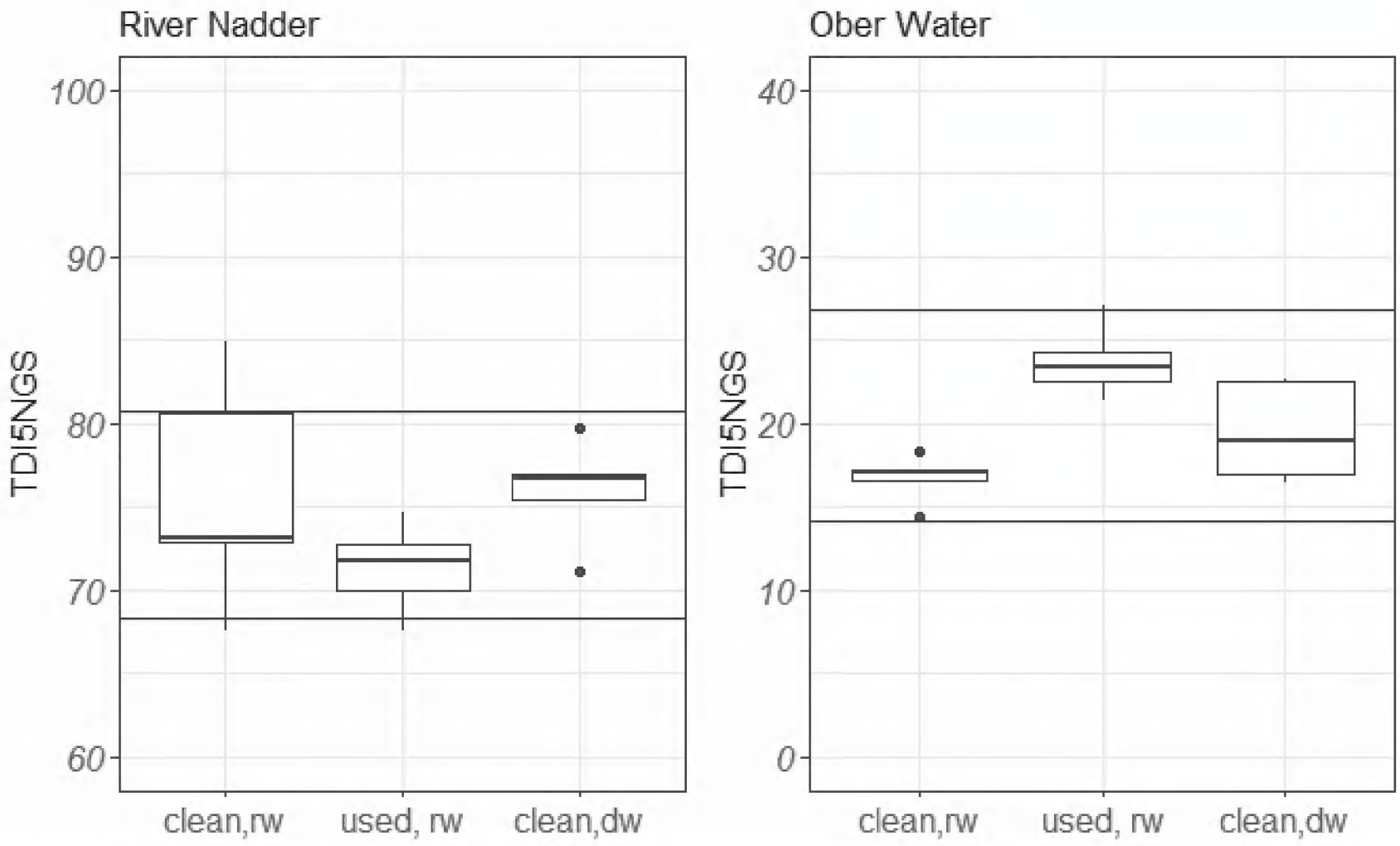

Figure 5. Variation in TDI5NGS between treatments for River Nadder and Ober Water. Horizontal lines show the upper and lower limits of variation expected for replicate samples from a site on a single day, using the samples collected using clean toothbrushes and distilled water as the benchmark.

taxa in the biofilm when stream water was used to wash the stones (Figure 4, Kruskal-Wallis tests: all $P > 0.3$).

There is no significant difference between treatments when TDI5NGS scores are calculated on samples from the River Nadder (Kruskal-Wallis test: $p = 0.14$). By contrast (and counterintuitively), TDI5NGS is significantly lower in samples from Ober Water which were removed with toothbrushes formerly used in the more enriched River Nadder (Figure 5: Kruskal-Wallis test: $p = 0.021$). Despite this, the scale of variation observed in each stream still lies within the range of variation expected to occur at a site on a single day.

Discussion

The results of this study highlight a potential for toothbrushes to retain traces of the diatom assemblage (and, presumably, other constituents of stream biofilms), even after the routine cleaning procedure (washing bristles vigorously in the stream and rubbing against waders: Kelly et al. 1998). The scale of this contamination is relatively low but is, nonetheless, present. Based on experience at a number of locations, sampling a thick biofilm where there are entangling filamentous algae and then using the same toothbrush at a subsequent site with a very thin biofilm is more likely to lead to problems than the reverse situation. Similarly, given that the TDI is based on a weighted average equation where taxa tolerant to nutrient enrichment have higher scores than those associated with low nutrients (Kelly et al. 2008b; Kelly et al. 2020), sampling a 'clean' site after a visit to a 'polluted' one is more likely to result in problems than the other way around.

Contamination from the stream water used to wash the samples appears to be less of a problem. In the case of the River Nadder, planktonic taxa dominated the suspended diatom assemblage. Planktonic taxa do not contribute to the TDI5NGS score and so should have no effect on the final index value, However, several of these have large cells with multiple chloroplasts and there may be issues when sampling coincides with a plankton bloom (see Vasselon et al. 2018) as large numbers of these may reduce the sequencing depth of the target benthic taxa. The risk is small, but should not be ruled out entirely.

However, our results also show that contamination, both from dirty equipment and upstream sources of DNA, do influence the composition of assemblages and, therefore, it is reasonable to assume that they may affect the final assessment in a few cases. Earlier studies, using morphological identification by light microscopy, had found variation of up to 7 TDI units between replicate samples collected on the same day (Kelly et al. 2008a), which far exceeds the significant differences observed between samples collected with "clean" and "dirty" toothbrushes in this study. Contamination from toothbrushes is, therefore, an additional source of variation that can and should be controlled, rather than a threat to the integrity of existing protocols. With this in mind, a pragmatic and precautionary approach for routine monitoring would be to use clean water wherever possible (tap water is used in the UK) along with a clean (but not necessarily new) toothbrush for each sample taken. At the end of each sampling trip, all toothbrushes should be cleaned in bleach or with hydrogen peroxide to avoid contamination on future occasions. A more stringent approach to contamination, however, may be appropriate in the future when data are not processed using the current generation of assessment tools, based on weighted average equations. Where the primary focus is the thorough documentation of diversity at a site, the introduction of even a small number of taxa may give misleading results. In such situations, a more stringent approach to contamination should be followed, with new equipment and distilled water used for each sample.

Acknowledgements

This research was funded by the Environment Agency of England: Project number SC160014.

References

Bailet B, Apothéoz-Perret-Gentil L, Baričević A, Chonova T, Franc A, Frigerio J-M, Kelly M, Mora D, Pfannkuchen M, Proft S, Ramon M, Vasselon V, Zimmermann J, Kahlert M (2020) Diatom DNA metabarcoding for ecological assessment: comparison among bioinformatics pipelines used in six European countries reveals the need for standardization. Science of the Total Environment 745: e140948. https://doi.org/10.1016/j.scitotenv.2020.140948

Bista I, Carvalho GR, Walsh K, Seymour M, Hajibabaei M, Lallias D, Christmas M, Creer S (2017) Annual time-series analysis of aqueous eDNA reveals ecologically relevant dynamics of lake ecosystem biodiversity. Nature Communications 8: article number 14087. https://doi.org/10.1038/ncomms14087

Bista I, Carvalho GR, Tang M, Walsh K, Zhou X, Hajibabaei M, Shokralla S, Seymour M, Bradley D, Liu S, Christmas M, Creer S (2018) Performance of amplicon and shotgun sequencing for accurate biomass estimation in invertebrate community samples. Molecular Ecology Resources 18: 1020–1034. https://doi.org/10.1111/1755-0998.12888

Blackman RC, Constable D, Hahn C, Sheard AM, Durkota J, Hänfling B, Handley LL (2017) Detection of a new non-native freshwater species by DNA metabarcoding of environmental samples – first record of Gammarus fossarum in the UK. Aquatic Invasions 12: 177–189. https://doi.org/10.3391/ai.2017.12.2.06

CEN (2014) EN 14407:2014. Water quality – Guidance standard for the identification, enumeration and interpretation of benthic diatom samples from running waters. Geneva: Comité European de Normalisation.

Charles DF, Kelly MG, Stevenson RJ, Poikane S, Theroux S, Zgrundo A, Cantonati M (2020) Benthic algal assessments in the EU and the US: striving for consistency in the face of great ecological diversity. Ecological Indicators 121: e107082. https://doi.org/10.1016/j.ecolind.2020.107082

Environment Agency (2018) A DNA based diatom metabarcoding approach for Water Framework Directive classification of rivers. Report SC140024/R. Environment Agency, Bristol.

Hänfling B, Handley LL, Read DS, Hahn C, Li J, Nichols P, Blackman RC, Oliver A, Winfield IJ (2016) Environmental DNA metabarcoding of lake fish communities reflects long-term data from established survey methods. Molecular Ecology. https://doi.org/10.1111/mec.13660

Kelly MG, Zgrundo A (2013) Potential for cross-contamination of benthic diatom samples when using toothbrushes. Diatom Research 28: 359–363. https://doi.org/10.1080/0269249X.2013.806959

Kelly MG, Cazaubon A, Coring E, Dell'Uomo A, Ector L, Goldsmith B, Guasch H, Hürlimann J, Jarlman A, Kawecka B, Kwandrans J, Laugaste R, Lindstrøm E-A, Leitao M, Marvan P, Padisák J, Pipp E, Prygiel J, Rott E, Sabater S, van Dam H, Vizinet J (1998) Recommendations for the routine sampling of diatoms for water quality assessments in Europe. Journal of Applied Phycology 10: 215–224. https://doi.org/10.1023/A:1008033201227

Kelly MG, Juggins S, Bennion H, Burgess A, Yallop M, Hirst H, King L, Jamieson BJ, Guthrie R, Rippey B (2008a) Use of diatoms for evaluating ecological status in UK freshwaters. Science Report: SC030103/SR4, Environment Agency, Bristol.

Kelly M, Juggins S, Guthrie R, Pritchard S, Jamieson J, Rippey B, Hirst H, Yallop M (2008b) Assessment of ecological status in U.K. rivers using diatoms. Freshwater Biology 53: 403–422. https://doi.org/10.1111/j.1365-2427.2007.01903.x

Kelly MG, Juggins S, Mann DG, Sato S, Glover R, Boonham N, Sapp M, Lewis E, Hany U, Kille P, Jones T, Walsh K (2020) Development of a novel metric for evaluating diatom assemblages in rivers using DNA metabarcoding. Ecological Indicators 118: e106725. https://doi.org/10.1016/j.ecolind.2020.106725

McCune B, Grace JB (2002) Analysis of ecological communities. MJM Software Design, Glenarden Beach, Oregon.

Muri D, Lawson Handley L, Bean CW, Li J, Peirson G, Sellars GS, Walsh K, Watson HV, Winfield IJ, Hänfling B (2020) Read counts from environmental DNA (eDNA) metabarcoding reflect fish abundance and biomass in drained ponds. Metabarcoidng and Metagenomics 4: 97–112. https://doi.org/10.3897/mbmg.4.56959

Oksanen J, Kindt R, Legendre P, O'Hara RB (2006) vegan: Community Ecology Package Version 1.8-5. http://CRAN.R-PROJECT.ORG/

R Development Core Team (2017) R: A Language and Environment for Statistical Computing. Reference Index. Version 3.4.1 (2017-06-30). Vienna: R Foundation for Statistical Computing. https://cran.r-project.org/manuals.html

Rimet F, Gusev E, Kahlert M, Kelly M, Kulikovskiy M, Maltsev Y, Mann D, Pfannkuchen M, Trobajo R, Vasselon V, Zimmermann J, Bouchez A (2019) Diat.barcode, an open-access curated barcode library for diatoms. Scientific Reports 9, article number: 15116. https://doi.org/10.1038/s41598-019-51500-6

Rimet F, Aylagas E, Borja A, Bouchez A, Canino A, Chauvin C, Chonova T, Ciampor Jr F, Costa FO, Ferrari BJD, Gastineau R, Goulon C, Gugger M, Holzmann M, Jahn R, Kahlert M, Kusber W-H, Laplace-Treyture C, Leese F, Leliaert F, Mann DG, Marchand F, Méléder V, Pawlowski J, Rasconi S, Rivera S, Rougerie R, Schweizer M, Trobajo R, Vasselon V, Vivien R, Weigand A, Witkowski A, Zimmermann J, Ekrem T (2021) Metadata standards and practical guidelines for specimen and DNA curation when building barcode reference libraries for aquatic life. Metabarcoding and Metagenomics 5: e58056. https://doi.org/10.3897/mbmg.5.58056

Vasselon V, Bouchez A, Rimet F, Jacquet S, Trobajo R, Corniquel M, Tapolczai K, Domaizon I (2018) Avoiding quantification bias in metabarcoding: application of a cell biovolume correction factor in diatom molecular biomonitoring. Methods in Ecology and Evolution 9: 1060–1069. https://doi.org/10.1111/2041-210X.12960

Vasselon V, Rimet F, Domaizon I, Monnier O, Reyjol Y, Bouchez A (2019) Assessing pollution of aquatic environments with diatoms' DNA metabarcoding: experience and developments from France Water Framework Directive networks. Metabarcoding and Metagenomics 3: e39646. https://doi.org/10.3897/mbmg.3.39646

Zelinka M, Marvan P (1961) Zur Prazisierung der biologischen Klassifikation des Reinheit fliessender Gewasser. Archiv für Hydrobiologie 57: 389–407.